Appunti di Misure Elettriche
Gli errori nelle misure

CLASSIFICAZIONE DEGLI ERRORI

Sappiamo che ogni misura è necessariamente affetta da un **errore**. Vogliamo adesso inquadrare le principali cause cui sono dovuti gli errori, al fine di capire se e in che modo essi possono essere ridotti al minimo. Sussiste allora la seguente classificazione:

- gli **errori grossolani** sono quelli addebitabili a imperizia o distrazione dell'operatore che sta compiendo la misura; possono ad esempio derivare da una sbagliata lettura o da un uso improprio degli strumenti, oppure da trascrizioni errate dai dati sperimentali o anche da errate elaborazioni di tali dati. E' evidente perciò che tali errori non si presentano quando si opera con cura e attenzione e comunque possono essere eliminati semplicemente ripetendo la misura;

- gli **errori sistematici** sono quelli che si ripresentano sempre con lo stesso segno e la stessa ampiezza, ove ovviamente la misura di una grandezza venga ripetuta più volte con la stessa strumentazione e nelle stesse condizioni operative ed ambientali. Questi errori sono generalmente dovuti ad una non corretta taratura degli strumenti oppure a difetti intrinseci degli strumenti stessi[1]; questo significa, sostanzialmente, che in talune situazioni è possibile correggere tali errori o comunque minimizzarli;

- infine, gli **errori accidentali** (detti anche *non sistematici* o *casuali*) sono quelli che permangono anche nell'ipotesi di essere riusciti ad eliminare tutti gli errori grossolani e sistematici. Le cause di tali errori sono tipicamente le imprevedibili fluttuazioni nelle condizioni operative, strumentali ed ambientali. Gli errori accidentali possono essere analizzati statisticamente, in quanto si è visto empiricamente che essi sono generalmente distribuiti secondo leggi semplici. In particolare, si fa spesso l'ipotesi che le cause di tali errori agiscano in modo del tutto aleatorio, determinando perciò scarti, rispetto al valore medio, sia negativi sia positivi. Questo ci autorizza ad aspettarci che gli effetti mediamente si annullino, ossia sostanzialmente che il *valore medio degli errori accidentali sia nullo*. Questa considerazione ha una conseguenza fondamentale: se riusciamo a correggere tutti gli errori grossolani e quelli sistematici, per cui avremo a che fare solo con errori accidentali, ci

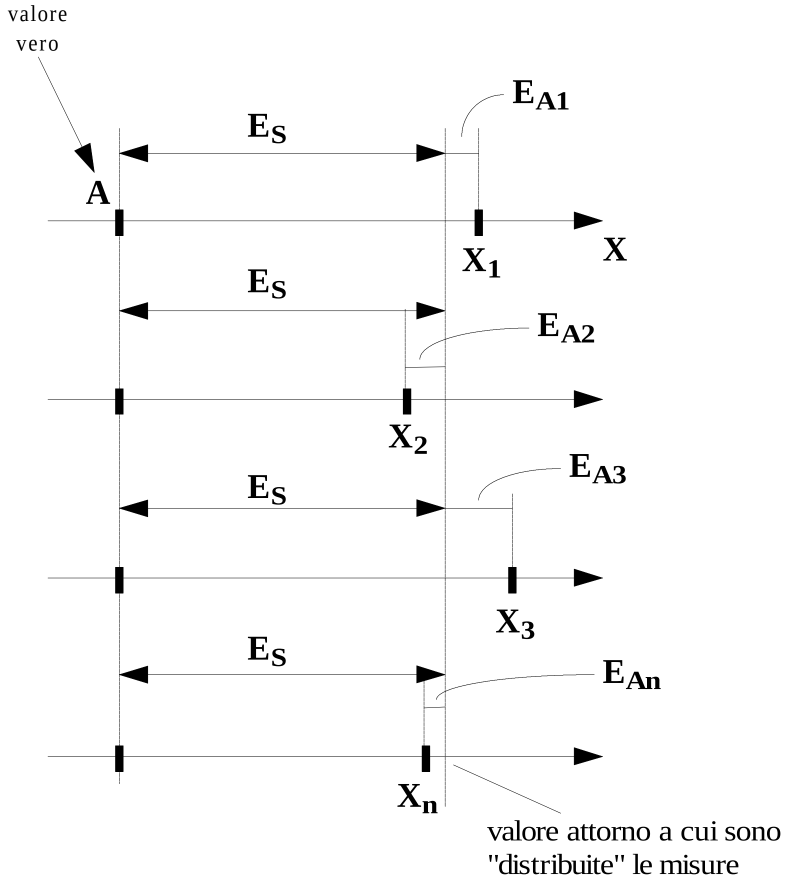
[1] Ad esempio, possono esserci dei difetti costruttivi oppure dei malfunzionamenti derivanti dall'aver usato gli strumenti in particolari condizioni operative o ambientali (elevate temperature, forti campi elettromagnetici, sovraccarichi e così via).

basterà compiere misure ripetute e poi mediare i risultati: quante più misure considereremo, tanto meno il risultato finale (media dei singoli risultati) sarà affetto da errori accidentali. *Quanto più piccoli risultano gli errori accidentali, tanto più si dice che la misura è **precisa***.

In generale, dunque, nell'ipotesi di aver eliminato ogni tipo di errore grossolano, possiamo affermare che *l'errore di una misura è somma di un errore sistematico (che si ripete ogni misura, in uguali condizioni operative) e di un errore accidentale (che invece varia casualmente in ogni misura, anche se le condizioni operative rimangono immutate):*

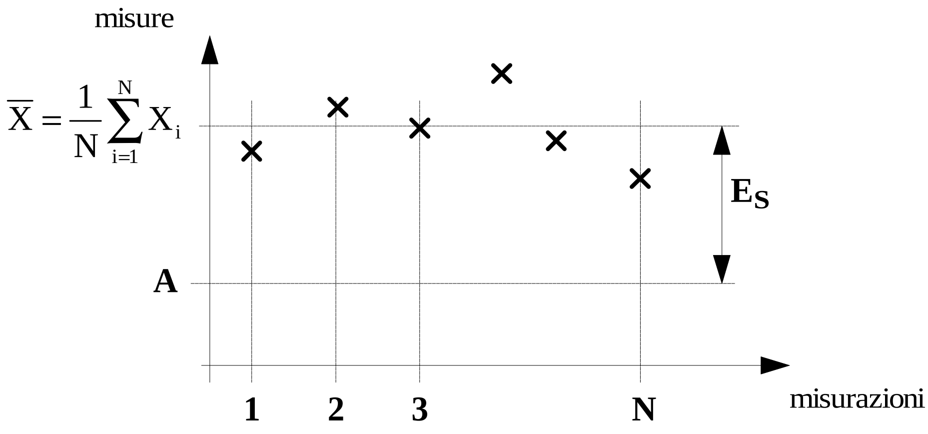
$$E = E_S + E_A$$

Il diagramma seguente aiuta a comprendere questi concetti:



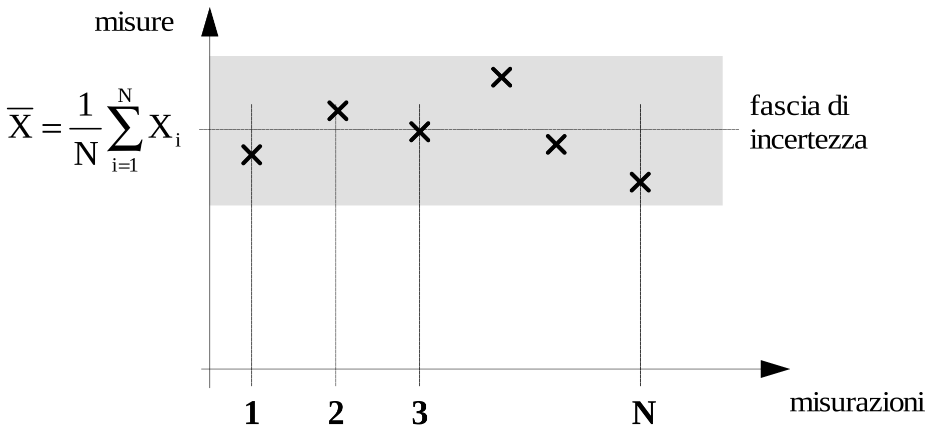
E' qui illustrato quello che accade quando vengono compiute N misure ripetute di una stessa quantità X, il cui valore vero (incognito) è A. Si nota che ciascuna misura X_i è affetta da un identico errore sistematico $\mathbf{E_S}$ (pari alla differenza tra A ed il valore attorno al quale risultano distribuite le X_i, ad esempio il loro valor medio), mentre ciò che cambia, di volta in volta, è l'errore accidentale $\mathbf{E_{Ai}}$, che rende diversa una misura dall'altra. Come si vede, l'errore accidentale cambia sia in ampiezza sia in segno.

Possiamo anche usare una rappresentazione grafica alternativa alla precedente. Usiamo in particolare un diagramma cartesiano che presenta in ascisse i numeri identificativi delle varie misurazioni (1,2,...,N) ed in ordinate i valori ottenuti per le misure:



Così come nel caso precedente, vengono qui riportati (questa volta in ordinate) il valore vero A (incognito) del misurando e i valori ottenuti dalle varie misure; tali valori sono "distribuiti" attorno al loro valor medio $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ (si veda in proposito quanto detto nei prossimi paragrafi), che differisce da A per una quantità pari, per definizione, all'errore sistematico[2]. La differenza tra il risultato X_i di ciascuna misura ed A è invece l'errore accidentale della misura i-sima, che cambia da caso a caso.

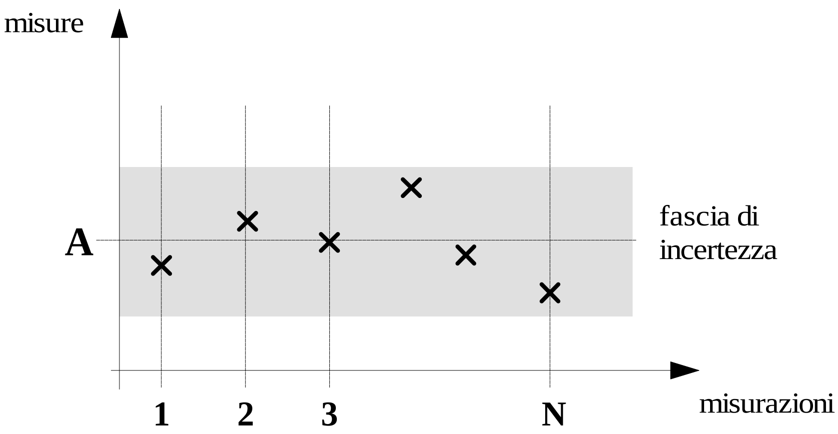
La dispersione dei valori X_i attorno al valor medio delimita una regione, detta **fascia di incertezza**, in cui risultano comprese tutte le misure effettuate:



Facciamo notare, comunque, che la ***fascia di incertezza*** *è definita rigorosamente come quella regione, a cavallo del valore vero A del misurando, nella quale si distribuiscono i risultati delle varie misure.* Tuttavia, il discorso fatto poco fa si adatta a questa definizione nell'ipotesi che il valore medio $\overline{X}$ si approssimi al valore vero A, il che accade quando **N**$\rightarrow\infty$, ossia nell'ipotesi di compiere un numero sufficientemente elevato di misure ripetute.

In modo del tutto equivalente, è evidente che, se riusciamo ad eliminare l'errore sistematico (il quale, come visto, incide in egual misura sulle X_i), le misure X_i saranno "distribuite" proprio attorno ad A e differiranno da essa per una quantità E_{Ai} (incognita) variabile da caso a caso:

[2] La figura riportata non deve trarre in inganno: i risultati X_i delle singole misure possono, in generale, risultare sia superiori sia inferiori ad A, anche se la figura riporta solo valori X_i superiori ad A.



La *caratteristica fondamentale degli errori accidentali è quella di essere a valor medio nullo,* per cui è nullo il valore medio degli scarti E_{Ai} di ciascuna misura rispetto ad A.

Errori sistematici

Soffermiamoci adesso maggiormente sugli **errori sistematici**, dovuti essenzialmente, come si è detto, a imperfezioni (costruttive o di taratura) degli strumenti impiegati per compiere le misure.

In primo luogo, è evidente che *è possibile ridurre gli errori sistematici effettuando una nuova e migliore **taratura** degli strumenti, usando tali strumenti nel modo appropriato e sottoponendoli ad una accurata e periodica **manutenzione**.*

In secondo luogo, gli errori sistematici dipendono anche dalle **condizioni ambientali**, ossia dall'ambiente in cui si esegue la misura: questo perché eventuali variazioni della *temperatura* oppure presenza di eventuali *campi elettromagnetici* sono fenomeni che possono influenzare sia la strumentazione sia lo stesso misurando. Si dice in questi casi che esiste una **interferenza esterna** sul sistema di misura e gli errori prendono anche il nome di **errori condizionati**.

In generale, diciamo che *gli errori sistematici sono difficili da valutare ed hanno anche una maggiore importanza di quelli accidentali:* infatti, mentre gli errori accidentali influenzano la **precisione** di una misura, gli errori sistematici influenzano l'**incertezza** della misura stessa.

Il modo classico di valutare l'errore sistematico su una misura è quello di confrontare una **stima A** nota del valore del misurando ed il **risultato X** della misura[3]:

$$\boxed{E_S = X - A}$$

Questa è in pratica la definizione classica di errore di una misura. Da notare che essa differisce dal modo con cui valutare gli **errori accidentali**: infatti, sfruttando il fatto che gli errori accidentali si possono ritenere a valor medio nullo, tali errori accidentali devono essere calcolati come differenza tra il risultato X di una singola misura e la media dei risultati di una serie di misure ripetute:

$$\boxed{E_A = X - \overline{X}}$$

[3] Al posto del risultato della singola misura, si potrebbe prendere il valor medio dei risultati di misure ripetute.

dove quindi $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$.

Tornando agli errori sistematici, *si definisce **correzione** il valore da aggiungere, algebricamente, al risultato non corretto di una misura per compensarne l'errore sistematico.* Il concetto è semplice: se il risultato X della misura è affetto da un errore sistematico $E_S = X - V$, allora potrò compensare tale errore sommando, ad X, la correzione **C=-E_S**:

$$X + C = X - E_S = X - (X - V) = V$$

Così facendo, otteniamo proprio il valore del misurando.

E' ovvio che il discorso è puramente teorico: infatti, non essendo perfettamente noto E_S (perché non è noto il valore V del misurando), non lo sarà nemmeno la correzione C e quindi la compensazione non sarà completa.

Un altro modo di esaminare lo stesso concetto consiste nel considerare il cosiddetto **fattore di correzione**, **C_F**, definito come quel numero per cui moltiplicare il risultato X di una misura al fine di compensare l'errore sistematico associato a tale misura: ciò significa porre

$$V = C_F X$$

Possiamo adesso fare la seguente considerazione: supponiamo di aver compiuto una data misura e di aver ottenuto il risultato X; questo risultato è affetto da un errore sistematico E_S che non è perfettamente noto; d'altra parte, esistono delle situazioni in cui è possibile ipotizzare un **modello matematico** dal quale ricavare l'entità dell'errore sistematico, il che ci consente una correzione dell'errore, come visto poco fa; il problema è che il modello ipotizzato non è mai perfetto (se lo fosse, potremmo eliminare l'errore), per cui non lo potrà essere anche la correzione; quindi, quando proviamo a correggere un errore sistematico, apportiamo una correzione anch'essa sicuramente imperfetta; di conseguenza, il risultato finale della nostra misura presenta due cause di incertezza: quella dovuta all'aver applicato una correzione non perfetta e quella associata alla presenza di effetti accidentali.

Quindi, dopo la correzione, il risultato potrebbe anche essere molto vicino al valore vero V del misurando, ossia potrebbe avere un errore sistematico residuo molto piccolo, ma potrebbe comunque essere affetto da grande incertezza, in quanto i fattori che la determinano non vanno confusi con gli errori. In altre parole, *è bene non confondere l'errore sistematico residuo non corretto con l'incertezza del risultato di una misura: pur riducendo l'errore sistematico, l'incertezza potrebbe rimanere elevata.*

ACCURATEZZA E PRECISIONE

Abbiamo capito, nei paragrafi precedenti, che ogni misura è soggetta a delle limitazioni, per cui il risultato della misura deve necessariamente essere accompagnato dall'indicazione quantitativa dell'**incertezza** della misura stessa.

Bisogna stare attenti a non confondere termini come accuratezza, incertezza e precisione. Per **accuratezza** (*accuracy*) intendiamo *il grado di approssimazione della quantità misurata al valore vero del misurando oppure ad una sua stima.*

Si tratta di un concetto tipicamente qualitativo e non quantitativo. Se invece si vuole passare ad una valutazione qualitativa, si considera l'**accuratezza relativa** (simbolo: **a**), definita nel modo seguente: se X è il risultato di una data misura ed E_S è l'errore sistematico (solo stimabile) di tale misura, l'accuratezza relativa è data da

$$\boxed{a = 1 - \left|\frac{E_S}{X}\right|}$$

E' evidente che, al diminuire dell'errore sistematico, aumenta l'accuratezza, fino al valore massimo 1 (teorico, non raggiungibile) quando $E_S=0$.

Ricordiamo che l'errore sistematico è definito come $E_S = X - V$, per cui possiamo anche scrivere che

$$a = 1 - \left|\frac{X - V}{X}\right| = 1 - \left|1 - \frac{V}{X}\right|$$

La quantità sotto il segno di valore assoluto è sicuramente minore di 1, per cui il valore minimo dell'accuratezza relativa è 0. Talvolta, si sente dire che uno strumento presenta una accuratezza dello 0.5%: se ci si riferisse all'accuratezza relativa appena definita, ciò significherebbe che lo strumento fornisce delle pessime prestazioni. Al contrario, probabilmente ci si riferiva all'*incertezza*: in questo caso, in modo del tutto qualitativo, si dovrebbe semplicemente dire che lo strumento presenta un'ottima accuratezza.

Distinto dall'accuratezza è il concetto della *precisione*: *per* ***precisione*** *di una misura intendiamo l'indicazione <u>numerica</u> dell'approssimazione di un insieme ripetuto di misure della stessa quantità al valore medio dell'insieme di misure.* Vediamo di spiegarci meglio: supponiamo di condurre una serie di misure X_i (i=1,2,...,) della stessa quantità; calcolando il valor medio di tale misure, scriviamo che

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Allora, quanto più le singole misure X_i si avvicinano al loro valor medio $\overline{X}$, tanto più diremo che la misura è <u>qualitativamente precisa</u>.

Analiticamente, la precisione della generica misura X_i sarà

$$p_i = 1 - \left|\frac{X_i - \overline{X}}{\overline{X}}\right|$$

La media aritmetica delle $\mathbf{p_i}$ rappresenta invece la precisione del campione di misura $(X_1,X_2,..,X_N)$:

$$p = \frac{1}{N}\sum_{i=1}^{N} p_i$$

Sottolineiamo che, pur potendo quantificare l'accuratezza e la precisione di una misura o di un insieme di misure ripetute, si tratta comunque di concetti essenzialmente qualitativi.

Concetto in qualche modo analogo a quello della precisione è quello della **ripetibilità**, definita come *il grado di approssimazione esistente tra i risultati di misure successive della stessa quantità, eseguite nelle <u>stesse condizioni di misura</u>.* Come vedremo più avanti, la deviazione standard delle misure eseguite rappresenta un indice di ripetibilità delle misure stesse.

Analogamente, *la cosiddetta **riproducibilità** è definita come il grado di approssimazione esistente tra i risultati di misure successive della stessa quantità, eseguite in diverse condizioni di misura.*

Sottolineiamo ancora una volta, come traspare dalle definizioni fornite, che l'accuratezza e la precisione sono concetti qualitativi e non quantitativi. Si può vedere concretamente che una bassa incertezza di misura si può avere solo quando sia l'accuratezza sia la precisione sono elevate. Accuratezza e precisione dipendono da vari fattori, tra cui citiamo la qualità degli strumenti utilizzati e la cura esercitata dall'utente nel compiere la misura.

Avere una elevata precisione significa avere sia una elevata ripetibilità delle misure sia un sufficiente numero di cifre significative: infatti, quanto maggiore è la precisione della misura tante più cifre significative la rappresentano e tanto più piccoli sono gli scarti tra una misura e l'altra. Al contrario, quando sono poche le cifre significative che rappresentano la misura, la precisione è piccola anche se gli scarti tra le varie misure sono piccoli: questo proprio perché tali scarti interessano cifre via via più significative.

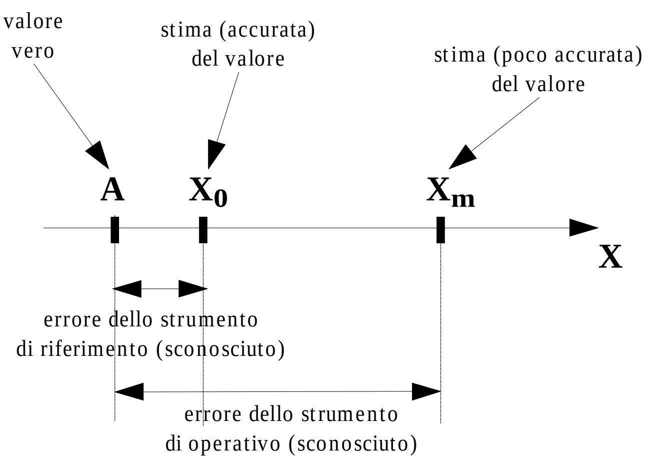
Facciamo un esempio: supponiamo di avere uno **strumento a lettura digitale** (ad esempio un *multimetro digitale*), in cui cioè il risultato della misura sia indicato su un display digitale con un certo numero di cifre; se, per esempio, le cifre fossero solo due, avremmo probabilmente una serie di misure ripetibili (cioè con risultati simili tra loro), ma non precise (data l'assenza di ulteriori cifre significative da leggere).

Viene ora da chiedersi se la precisione implica l'accuratezza e/o viceversa. Si può affermare che *la precisione è una condizione necessaria ma non sufficiente per l'accuratezza:* questo significa che una misura, per essere accurata, deve anche essere precisa (e quindi rappresentabile con un elevato numero di cifre significative), ma una misura precisa potrebbe comunque essere non accurata. Facciamo anche qui un esempio: supponiamo di avere uno strumento che permetta di leggere 6 cifre della grandezza da misurare e supponiamo inoltre di compiere diverse misure, abbastanza ravvicinate (nel tempo) tra di loro; in queste condizioni, possiamo sicuramente affermare di aver compiuto una misura precisa se le misure si scostano poco una dall'altra, ma non è detto che la misura sia accurata, ad esempio perché lo strumento non è ben tarato.

Fin qui, dunque, l'accuratezza e la precisione ci forniscono indicazioni solo qualitative sulla bontà delle nostre misure. Se vogliamo delle indicazioni più quantitative, dobbiamo rivolgerci all'**incertezza**, la quale, come già detto, rappresenta sostanzialmente la nostra impossibilità a valutare con esattezza il misurando in questione. I prossimi paragrafi introducono alcuni concetti di statistica utili per comprendere il modo con cui valutare l'incertezza.

ERRORE STIMATO

Supponiamo di voler misurare una certa grandezza X, ad esempio una resistenza, e supponiamo che sia A il valore vero (a noi sconosciuto) di tale quantità. Supponiamo inoltre di compiere due distinte misure di X, una con uno strumento molto accurato (che chiamiamo **strumento di riferimento**) ed una con uno strumento meno accurato (detto **strumento operativo**), ad esempio quello usato in un laboratorio universitario. I due strumenti ci daranno due distinte misure, che indichiamo rispettivamente con X_0 ed X_m. Possiamo disporre tali valori, insieme al valore vero A, su un asse orizzontale, al fine di dare una interpretazione grafica di ciò che stiamo facendo:



La quantità $\mathbf{E_0=X_0-A}$ rappresenta l'errore compiuto dallo strumento di riferimento: si tratta di una quantità sconosciuta (dato che non si conosce A), ma comunque piccola, data la bontà dello strumento usato.

Analogamente, la quantità $\mathbf{E_m=X_m-A}$ rappresenta l'errore compiuto dallo strumento operativo: anche questa è una quantità sconosciuta e si presume inoltre più grande di E_0, data la minore bontà dello strumento usato.

L'unica quantità nota è la differenza E_m-E_0 tra l'errore dello strumento operativo e quello dello strumento di riferimento: infatti, possiamo scrivere che

$$E_m - E_0 = (X_m - A) - (X_0 - A) = X_m - X_0$$

Questa quantità può allora essere presa come una **stima dell'errore dello strumento operativo**. In altre parole, quindi, *se vogliamo "stimare" l'errore commesso dal nostro strumento operativo, dobbiamo necessariamente fare affidamento ad uno strumento di riferimento (o, quanto meno, ad un campione).*

MEDIA, DEVIAZIONE MEDIA, DEVIAZIONE STANDARD E VARIANZA DI UN CAMPIONE DI MISURA

Supponiamo di compiere N misure successive dello stesso misurando; se $\mathbf{X_i}$ è il risultato della generica misura, il nostro **campione di misura** è definito come la N-pla delle X_i:

$$(X_1, X_2, \ldots, X_N)$$

Tramite gli elementi di questo vettore possiamo calcolare una serie di quantità. La principale di esse è la **media aritmetica** delle N misure, già introdotta precedentemente:

$$\boxed{\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i}$$

Come è abbastanza intuitivo aspettarsi, questa quantità rappresenta la <u>migliore stima possibile</u> che possiamo fornire del nostro misurando.

In taluni casi, del resto, ha senso calcolare, al posto della media aritmetica, una **media pesata**, in modo da attribuire maggior peso ad alcune misure rispetto alle altre. Si considera perciò la seguente quantità:

$$\boxed{\overline{X}_p = \frac{\sum_{i=1}^{N} w_i X_i}{\sum_{i=1}^{N} w_i}}$$

Si tratta ovviamente di definire nel modo più opportuno possibile i **pesi** $\mathbf{w_i}$ secondo cui effettuare tale media pesata. I criteri possono essere diversi: ad esempio, si potrebbe dare maggior peso alle misure che sono più attendibili o comunque maggiormente significative. Il criterio più sensato si basa sulle seguenti considerazioni: dobbiamo infatti considerare che ciascuna misura X_i è affetta da una propria **incertezza** (che indichiamo con $\mathbf{u_i}$); allora, *si può attribuire maggior peso alle misure affette da incertezza minore;* per ottenere questo, bisogna semplicemente prendere dei coefficienti di peso che siano inversamente proporzionali all'incertezza delle corrispondenti misure: generalmente, si pone allora

$$w_i = \frac{1}{2u_i^2}$$

Notiamo che, ovviamente, se le misure avessero tutte la stessa incertezza, i coefficienti di peso sarebbe uguali e quindi la media pesata risulterebbe pari alla media aritmetica.

Consideriamo adesso la media aritmetica[4] e facciamo qualche passaggio in più. In particolare, indichiamo, rispettivamente, con E_{si} e E_{ai} gli errori sistematici e accidentali sulla i-sima misura X_i; allora, quest'ultima può essere scritta come

$$X_i = V + E_i = V + E_{si} + E_{ai}$$

dove ovviamente la quantità V (valore vero del misurando) non è affetta da indice in quanto le misure sono tutte relative allo stesso misurando.

Calcolando adesso la media aritmetica dei due membri, a primo membro otteniamo la media aritmetica delle N misure, mentre a secondo membro otteniamo quanto segue:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} V + \frac{1}{N}\sum_{i=1}^{N} E_{si} + \frac{1}{N}\sum_{i=1}^{N} E_{ai} = V + \frac{1}{N}\sum_{i=1}^{N} E_{si} + \frac{1}{N}\sum_{i=1}^{N} E_{ai}$$

In base a quanto in precedenza, *gli errori accidentali rappresentano una tipica variabile aleatoria a valor medio zero;* questo vale, però, solo per $N \to \infty$, cioè per un numero molto grande di misure. Facendo allora questa ipotesi di N molto grande, possiamo concludere che

$$\overline{X} = V + \frac{1}{N}\sum_{i=1}^{N} E_{si}$$

Questa relazione ci suggerisce diverse considerazioni:

[4] Ricordiamo che la media aritmetica ci consente di valutare la ripetibilità delle misure effettuate, ossia il grado di approssimazione di tali misure al valore della loro media.

- in primo luogo, essa dice che la media aritmetica di un insieme di misure è sostanzialmente una stima del valore V del misurando, tanto migliore quanto minore è il valor medio degli errori sistematici sulle singole misure;
- in secondo luogo, se scriviamo quella relazione come

$$\text{bias} = \frac{1}{N}\sum_{i=1}^{N} E_{si} = \overline{X} - V$$

otteniamo la definizione del cosiddetto **bias** (*polarizzazione*) del nostro insieme di misure: esso è definito come il valor medio degli errori sistematici ed è quindi pari alla differenza tra il valor medio delle misure ed il valore vero del misurando. Ovviamente, il bias cambiato di segno rappresenta la correzione totale da apportare ad $\overline{X}$ per ottenere un miglioramento dell'accuratezza.

Si definisce inoltre **deviazione** della misura i-sima la differenza tra il risultato della misura e la media dei risultati delle N misure:

$$d_i = X_i - \overline{X}$$

Questo concetto consente di definire anche la **dispersione** dell'insieme di misure rispetto al valor medio. Ad esempio, un modo di quantificare tale dispersione è quello di calcolare il valor medio dei moduli delle deviazioni delle singole misure, ossia la cosiddetta **deviazione media**:

$$\alpha = \frac{1}{N}\sum_{i=1}^{N} |d_i| = \frac{1}{N}\sum_{i=1}^{N} |X_i - \overline{X}|$$

Il motivo per cui si considerano le deviazioni in modulo è semplicemente quello per cui tali deviazioni hanno valor medio nullo:

$$\frac{1}{N}\sum_{i=1}^{N} d_i = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}\right) = \frac{1}{N}\sum_{i=1}^{N} X_i - \frac{1}{N}\sum_{i=1}^{N}\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i - \overline{X} = 0$$

Ad ogni modo, il parametro tipicamente utilizzato per valutare la dispersione delle misure è la **deviazione standard**, definita in termini dei quadrati delle deviazioni delle singole misure, nel modo seguente:

$$\sigma \cong \sqrt{\frac{1}{N}\sum_{i=1}^{N} d_i^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}$$

Facciamo subito osservare che il segno di *all'incirca uguale* in quest'ultima relazione sarà chiarito più avanti.

Il quadrato della deviazione standard è la cosiddetta **varianza** del campione di misura:

$$\sigma^2 \cong \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i^2 + \overline{X}^2 - 2X_i\overline{X}\right) = \frac{1}{N}\sum_{i=1}^{N} X_i^2 + \frac{1}{N}\sum_{i=1}^{N}\overline{X}^2 - \frac{1}{N}\sum_{i=1}^{N} 2X_i\overline{X} =$$

$$= \frac{1}{N}\sum_{i=1}^{N} X_i^2 + \overline{X}^2 - 2\overline{X}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \overline{X}^2$$

Come ultima definizione, forniamo quella di **momento centrale di ordine q** del generico campione di misura:

$$E\left[X^q\right] \cong \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^q$$

Si vede subito che questa quantità è una generalizzazione di quelle introdotte in precedenza: infatti, per q=1 otteniamo una quantità nulla, mentre per q=2 otteniamo la varianza.

A conclusione del paragrafo, *sottolineiamo che le definizioni date in questo paragrafo sono riferite sempre al campione di misura che abbiamo a disposizione.*

CONCETTI DI FREQUENZA E DI CLASSI

Spesso, la comprensione di un fenomeno fisico può essere facilitata da un esame visivo dei risultati di misure ripetute di una grandezza o, comunque, in generale, di dati statistici. Bisogna allora individuare il modo migliore per rappresentare graficamente i dati disponibili. Ci vengono allora in aiuto alcuni concetti.

Sia dato il nostro campione di misure (X_1,X_2,....,X_N) composto da N elementi. Sia X_i il risultato della i-sima misura; è possibile che questo stessi risultato si ottenga in più di una misura; allora, il numero di volte $\mathbf{n_i}$ in cui la misura ha fornito valore X_i prende il nome di **frequenza** della misura X_i. La quantità $\mathbf{f_i=n_i/n}$ prende invece il nome di **frequenza relativa** di X_i.

Ovviamente, se tutte le misure forniscono risultati diversi, la frequenza delle singole misure sarebbe 1 e quindi le frequenze relative sarebbero tutte pari ad 1/n. Al contrario, se tutte le misure dessero lo stesso risultato, allora risulterebbe n_i=n e quindi f_i=1.

Un modo alternativo di definire la frequenza relativa si basa sul raggruppamento delle misure in gruppi (dette **classi**): si tratta banalmente di intervallini in cui le misure possono cadere, definiti perciò ciascuno da un estremo inferiore ed uno superiore. Se K sono in totale le classi individuate, l'ampiezza della i-sima classe sarà

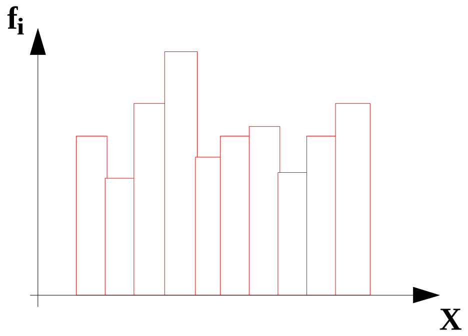
$$\Delta X_i = \frac{X_{max,i} - X_{min,i}}{K}$$

dove ovviamente $X_{max,i}$ e $X_{min,i}$ sono, rispettivamente, l'estremo superiore ed inferiore della classe i-sima.

Una volta effettuata questa suddivisione, diremo che la frequenza relativa della i-sima classe è il numero N_i di misure che cadono in tale classe, rapportato al numero totale N di misure:

$$f_i = \frac{N_i}{N}$$

Si può dare una interpretazione grafica. Consideriamo infatti un grafico cartesiano; in ascisse mettiamo il valore X delle misure, mentre in ordinate mettiamo la frequenza relativa f_i:



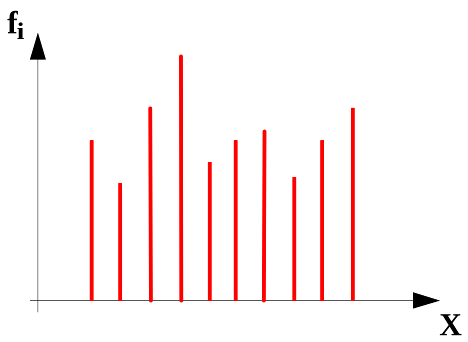
L'asse delle ascisse è stato suddiviso nelle varie classi, supposte tutte per comodità di ampiezza ΔX_i unitaria, il che significa che $X_{max,i} - X_{min,i} = K$, ossia che $X_{max,i} = X_{min,i} + K$.

Con questa scelta, otteniamo un istogramma, in cui la frequenza relativa rappresenta l'area del generico rettangolino.

Ovviamente, dato che la somma delle frequenze relative è pari ad 1, l'area complessiva sottesa dall'istogramma è a sua volta unitaria:

$$\sum_{i=1}^{N} f_i = \sum_{i=1}^{N} \frac{N_i}{N} = \frac{1}{N}\sum_{i=1}^{N} N_i = \frac{1}{N} N = 1$$

Ovviamente, ci sono da considerare due possibilità, a seconda dei risultati che possiamo attenderci dalle misure: il caso rappresentato nell'ultima figura è quello in cui le misure possono assumere qualsiasi valore reale compreso tra il minimo ed il massimo; l'altro caso, rappresentato nella figura seguente, è invece quello in cui i risultati possono assumere solo valori discreti (pensiamo ad esempio al lancio di due dadi, che può dare, come risultato, solo i numeri 2,3,....,12):



Per identificare ciascuna classe, possiamo considerare il valor medio delle misure che cadono in tale classe: possiamo cioè associare alla i-sima classe la quantità

$$X_i = \frac{1}{F_i}\sum_{h=1}^{F_i} X_{hi}$$

dove $\mathbf{F_i}$ è la frequenza della classe i-sima, ossia il numero di misure che cadono in tale classe, e dove $\mathbf{X_{hi}}$ è la h-sima misura che cade nella classe i-sima.

Adesso, possiamo calcolare nuovamente la media delle N misure:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{K}\sum_{h=1}^{F_i} X_{hi} = \frac{1}{N}\sum_{i=1}^{K} F_i\left(\frac{1}{F_i}\sum_{h=1}^{F_i} X_{hi}\right) = \frac{1}{N}\sum_{i=1}^{K} F_i X_i = \frac{1}{N}\sum_{i=1}^{K} N f_i X_i = \sum_{i=1}^{K} f_i X_i$$

dove abbiamo evidentemente tenuto conto che frequenza e frequenza relativa sono legate dalla relazione $f_i=F_i/N$.

La formula appena ricavata mette in evidenza che la media aritmetica delle misure non dipende più dal numero N di prove eseguite: notiamo infatti che essa è pari ad una media pesata dei valori medi X_i delle singole classi considerate, dove i coefficienti di peso sono le frequenze relative associate alle classi.

Naturalmente, basandoci sulle X_i, possiamo ripetere gli stessi discorsi fatti in precedenza circa la deviazione media, la deviazione standard e la varianza, che assumono le seguenti espressioni (ottenute con procedimento analogo a quello seguito per il calcolo della media $\overline{X}$):

- deviazione media: $\alpha \cong \sum_{i=1}^{K} f_i \left|X_i - \overline{X}\right|$
- deviazione standard: $\sigma \cong \sqrt{\sum_{i=1}^{K} f_i \left(X_i - \overline{X}\right)^2}$
- varianza: $\sigma^2 \cong \sum_{i=1}^{K} f_i \left(X_i - \overline{X}\right)^2$

Autore: **SANDRO PETRIZZELLI**
e-mail: sandry@iol.it
sito personale: http://users.iol.it/sandry
succursale: http://digilander.iol.it/sandry1